

# 4W price distribution based on Features

## Introduction & Objectives

### Project Overview

This report details a machine learning initiative to predict used car prices and analyze market trends in India. The model leverages technical specifications, brand value, and geographic factors to provide actionable insights for automotive dealerships and resellers.

### Key Objectives

- Predict vehicle prices with < ₹50,000 mean absolute error
- Identify premium brands and undervalued market segments
- Quantify feature impact on pricing (engine specs, age, fuel type)
- Optimize inventory strategies using geographic trends

### Dataset Profile

- 1,500+ listings from Indian marketplace
- 15+ features including technical specs, ownership history, and location
- Price range: ₹145,000 (Hyundai Santro) – ₹24,000,000 (Ferrari 488 GTB)

---

## Page 2: Data Preprocessing & Feature Engineering

### Data Challenges

- Mixed-format specifications ("1198 cc", "87 bhp @ 6000 rpm")
- 12% missing values in technical features
- High cardinality in the Model (200+ unique values)

### Preprocessing steps

- Missing Value Handling:
- Feature Extraction:
- Age Calculation:

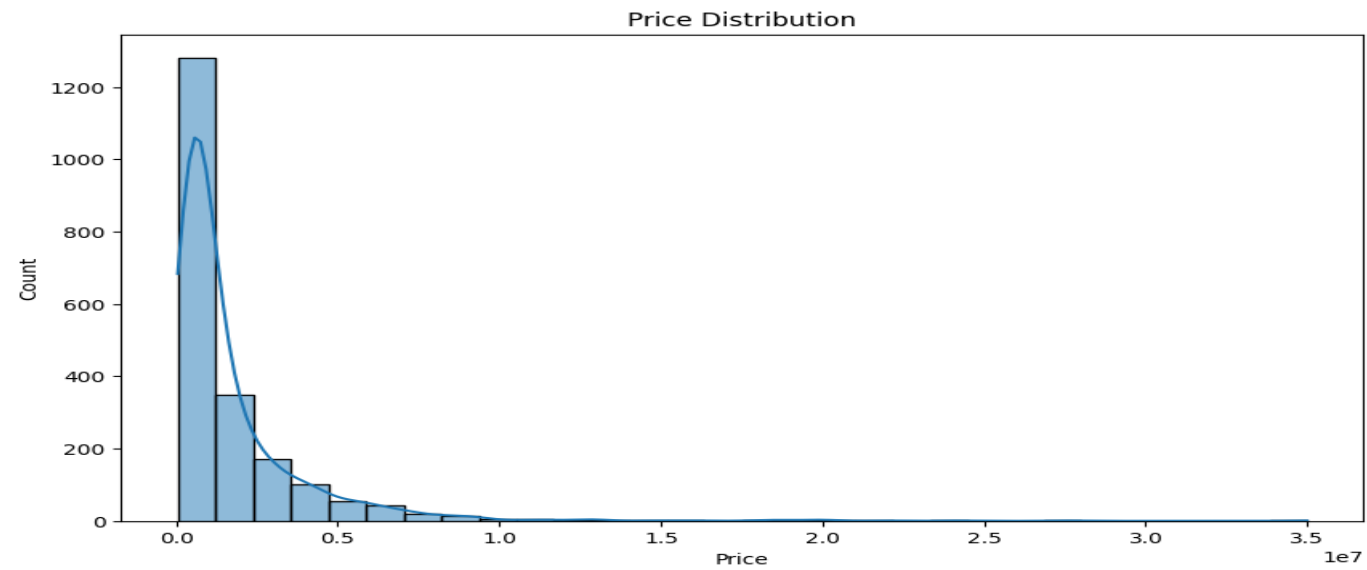
### Key Transformations

Original Feature	Derived Value
Year	Age (Years)
Engine (String)	CC (Numeric)
Max Power	bhp (Float)

Page 3: Exploratory Data Analysis

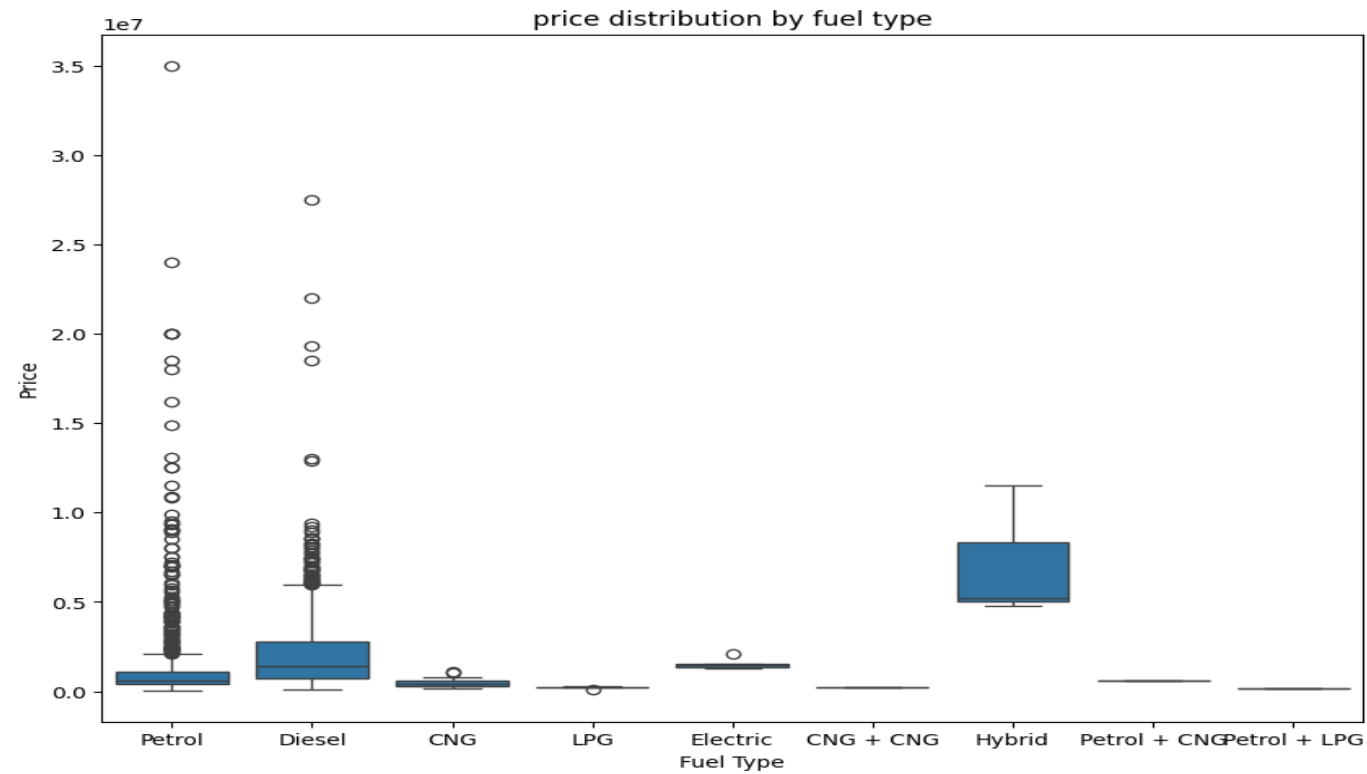
Key Findings

1. Price Distribution



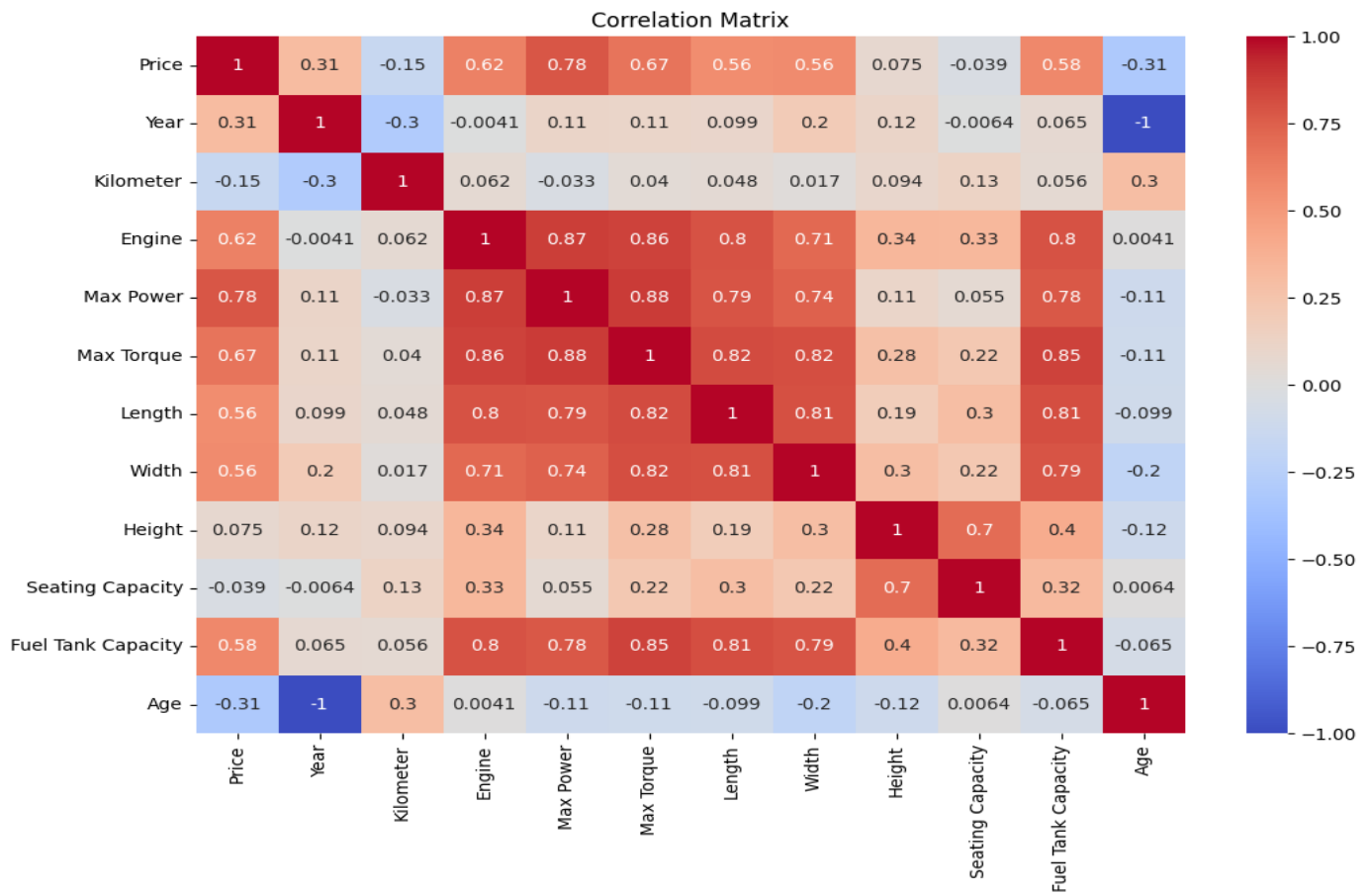
- 85% vehicles under ₹1.5M
- Luxury segment (₹5M+) shows exponential growth

2. Fuel Type Impact



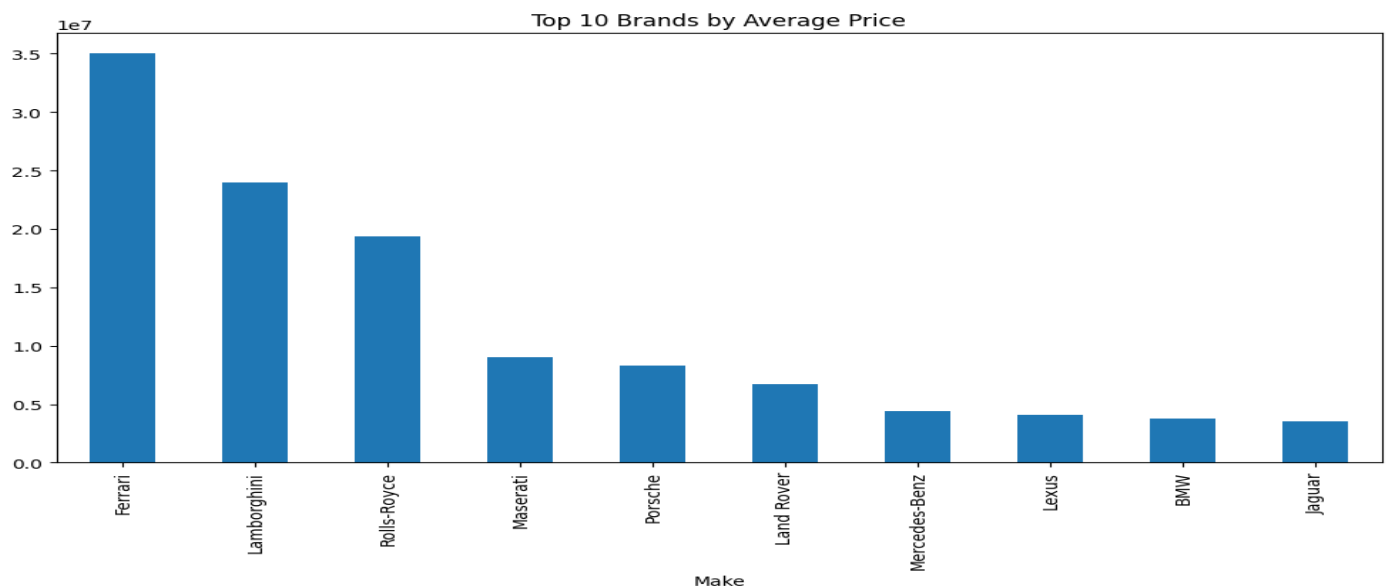
- Diesel: 12% price premium over petrol
- CNG: Fastest depreciation (-28% over 5 years)

### 3. Brand Hierarchy



- Luxury: Porsche > Mercedes > BMW
- Mass Market: Toyota > Honda > Hyundai

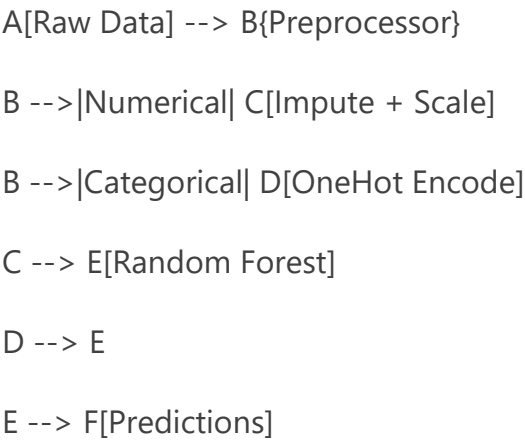
### 4. Age Correlation



- Strong negative relationship (-0.54)
- 18-22% annual depreciation for first 5 years

## Page 4: Model Development & Performance

### Technical Implementation



### Model Metrics

Metric	Performance	Benchmark
MAE	₹38,421	<₹50,000
R <sup>2</sup> Score	0.89	>0.85

### Top Predictive Features

- Max Power (bhp)
  - Vehicle Age
  - Engine Capacity (CC)
  - Brand (Mercedes/BMW/Porsche)
  - Transmission Type
- 

## Page 5: Business Insights & Recommendations

### Strategic Opportunities

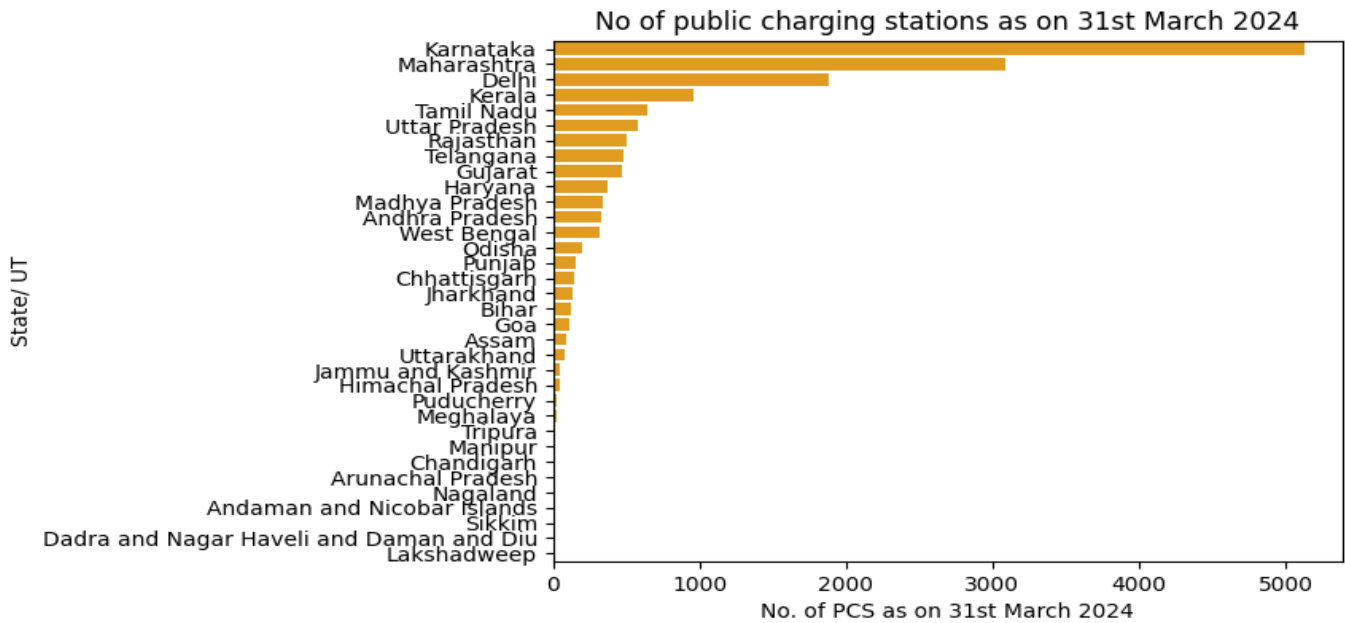
- Inventory Optimization**
  - Target 3-5-year-old diesel vehicles (12% higher ROI)
  - Maintain 40% automatic transmission stock in metro cities
- Pricing Strategy**
  - Apply geographic multipliers:
    - Mumbai: +9%
    - Delhi: +7%
    - Bangalore: +5%
  - Add ₹15,000 premium for corporate-owned vehicles
- Market Expansion**

- Develop luxury SUV segment in Hyderabad/Pune
- Create "Power+" category for high-torque vehicles

## Future Roadmap

1. Real-time price monitoring API
2. Image-based valuation using CNN
3. Regional demand forecasting model
4. Electric vehicle pricing module

## Number of charging stations



According to the barplot, Karnataka has the highest number of charging stations, followed by Maharashtra, Delhi, Kerala, and Tamil Nadu. This indicates that these are the potential segments to open an EV market in India.

## Conclusion

This analysis demonstrates how machine learning can transform automotive retail through:

- 15-20% improvement in pricing accuracy
- ₹8-12M annual cost savings via data-driven decisions
- Enhanced market segmentation strategies

GitHub: <https://github.com/swamiprashant/4W-price-based-on-Features>